

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/836,169DATE: 11/20/2001
TIME: 19:31:11

INPUT SET: S36674.raw

OIPE

#2

This Raw Listing contains the General
Information Section and up to the first 5 pages.

SEQUENCE LISTING

(1) General Information:

(i) APPLICANT: Choulika, Andre
Perrin, Arnaud
Dujon, Bernard
Nicolas, Jean-Francois

(ii) TITLE OF INVENTION: Nucleotide Sequence Encoding the Enzyme
I-SCEI and the Uses Thereof

(iii) NUMBER OF SEQUENCES: 52

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
Dunner

(B) STREET: 1300 I Street, N.W.

(C) CITY: Washington

(D) STATE: D.C.

(E) COUNTRY: USA

(F) ZIP: 20005-3315

ENTERED

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 09/836,169

(B) FILING DATE: 04-APRIL-2001

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/417,226

(B) FILING DATE: 05-APR-1995

(C) CLASSIFICATION:

(viii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/971,160

(B) FILING DATE: 05-NOV-1992

(ix) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/879,689

(B) FILING DATE: 05-MAY-1992

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/836,169DATE: 11/20/2001
TIME: 19:31:12

INPUT SET: S36674.raw

47
48 (x) ATTORNEY/AGENT INFORMATION:
49 (A) NAME: Potter, Jane E.R.
50 (B) REGISTRATION NUMBER: 33,332
51 (C) REFERENCE/DOCKET NUMBER: 03495-0111-04000
52
53 (xi) TELECOMMUNICATION INFORMATION:
54 (A) TELEPHONE: 202-408-4000
55 (B) TELEFAX: 202-408-4400
56
57 (2) INFORMATION FOR SEQ ID NO:1:
58
59 (i) SEQUENCE CHARACTERISTICS:
60 (A) LENGTH: 714 base pairs
61 (B) TYPE: nucleic acid
62 (C) STRANDEDNESS: single
63 (D) TOPOLOGY: linear
64
65 (ii) MOLECULE TYPE: DNA (genomic)
66
67
68 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
69
70 ATGCATATGA AAAACATCAA AAAAAACCAG GTAATGAACC TCGGTCCGAA CTCTAAACTG 60
71
72 CTGAAAGAAT ACAAATCCCA GCTGATCGAA CTGAACATCG AACAGTTCGA AGCAGGTATC 120
73
74 GGTCTGATCC TGGGTGATGC TTACATCCGT TCTCGTGATG AAGGTAAAAC CTACTGTATG 180
75
76 CAGTTCGAGT GGAAAAACAA AGCATACATG GACCACGTAT GTCTGCTGTA CGATCAGTGG 240
77
78 GTACTGTCCC CGCCGCACAA AAAAGAACGT GTTAACCACC TGGGTAACCT GGTAATCACC 300
79
80 TGGGGCGCCC AGACTTTCAA ACACCAAGCT TTCAACAAAC TGGCTAACCT GTTCATCGTT 360
81
82 AACAAACAAA AAACCATCCC GAACAACCTG GTTGAAAACT ACCTGACCCC GATGTCTCTG 420
83
84 GCATACTGGT TCATGGATGA TGGTGGTAAA TGGGATTACA AAAAAAACTC TACCAACAAA 480
85
86 TCGATCGTAC TGAACACCCA GTCTTTCACT TTCGAAGAAG TAGAATACCT GGTTAAGGGT 540
87
88 CTGCGTAACA AATTCCAACCT GAACTGTTAC GTAAAAATCA AAAAAACAA ACCGATCATC 600
89
90 TACATCGATT CTATGTCTTA CCTGATCTTC TACAACCTGA TCAAACCGTA CCTGATCCCCG 660
91
92 CAGATGATGT ACAAACAGCC GAACACTATC TCCTCCGAAA CTTTCCTGAA ATAA 714
93
94 (2) INFORMATION FOR SEQ ID NO:2:
95
96 (i) SEQUENCE CHARACTERISTICS:
97 (A) LENGTH: 237 amino acids
98 (B) TYPE: amino acid
99 (D) TOPOLOGY: linear

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/836,169DATE: 11/20/2001
TIME: 19:31:12

INPUT SET: S36674.raw

100
101 (ii) MOLECULE TYPE: peptide
102
103
104
105
106
107 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
108
109 Met His Met Lys Asn Ile Lys Lys Asn Gln Val Met Asn Leu Gly Pro
110 1 5 10 15
111
112 Asn Ser Lys Leu Leu Lys Glu Tyr Lys Ser Gln Leu Ile Glu Leu Asn
113 20 25 30
114
115 Ile Glu Gln Phe Glu Ala Gly Ile Gly Leu Ile Leu Gly Asp Ala Tyr
116 35 40 45
117
118 Ile Arg Ser Arg Asp Glu Gly Lys Thr Tyr Cys Met Gln Phe Glu Trp
119 50 55 60
120
121 Lys Asn Lys Ala Tyr Met Asp His Val Cys Leu Leu Tyr Asp Gln Trp
122 65 70 75 80
123
124 Val Leu Ser Pro Pro His Lys Lys Glu Arg Val Asn His Leu Gly Asn
125 85 90 95
126
127 Leu Val Ile Thr Trp Gly Ala Gln Thr Phe Lys His Gln Ala Phe Asn
128 100 105 110
129
130 Lys Leu Ala Asn Leu Phe Ile Val Asn Asn Lys Lys Thr Ile Pro Asn
131 115 120 125
132
133 Asn Leu Val Glu Asn Tyr Leu Thr Pro Met Ser Leu Ala Tyr Trp Phe
134 130 135 140
135
136 Met Asp Asp Gly Gly Lys Trp Asp Tyr Asn Lys Asn Ser Thr Asn Lys
137 145 150 155 160
138
139 Ser Ile Val Leu Asn Thr Gln Ser Phe Thr Phe Glu Glu Val Glu Tyr
140 165 170 175
141
142 Leu Val Lys Gly Leu Arg Asn Lys Phe Gln Leu Asn Cys Tyr Val Lys
143 180 185 190
144
145 Ile Asn Lys Asn Lys Pro Ile Ile Tyr Ile Asp Ser Met Ser Tyr Leu
146 195 200 205
147
148 Ile Phe Tyr Asn Leu Ile Lys Pro Tyr Leu Ile Pro Gln Met Met Tyr
149 210 215 220
150
151 Lys Leu Pro Asn Thr Ile Ser Ser Glu Thr Phe Leu Lys
152 225 230 235

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/836,169DATE: 11/20/2001
TIME: 19:31:12

INPUT SET: S36674.raw

153

154 (2) INFORMATION FOR SEQ ID NO:3:

155

156 (i) SEQUENCE CHARACTERISTICS:

157 (A) LENGTH: 722 base pairs

158 (B) TYPE: nucleic acid

159 (C) STRANDEDNESS: single

160 (D) TOPOLOGY: linear

161

162 (ii) MOLECULE TYPE: DNA (genomic)

163

164

165

166 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

167

168 AAAAATAAAA TCATATGAAA AATATTAAAA AAAATCAAGT AATCAATCTC GGCCTATTTT 60

169

170 CTAAATTATT AAAAGAATAT AAATCACAAT TAATTGAATT AAATATTGAA CAATTTGAAG 120

171

172 CAGGTATTGG TTTAATTTTA GGAGATGCTT ATATTCGTAG TCGTGATGAA GGTAAACTT 180

173

174 ATTGTATGCA ATTTGAGTGG AAAAATAAGG CACATGGA TCATGTATGT TTATTATATG 240

175

176 ATCAATGGGT ATTATCACCT CCTCATAAAA AAGAAAGAGT TAATCATTTA GGTAATTTAG 300

177

178 TAATTACCTG GGGAGCTCAA ACTTTTAAAC ATCAAGCTTT TAATAAATTA GCTAACTTAT 360

179

180 TTATTGTAAA TAATAAAAAA CTTATTCCTA ATAATTTAGT TGAAAATTAT TTAACACCTA 420

181

182 TGAGTCTGGC ATATTGGTTT ATGGATGATG GAGGTAAATG GGATTATAAT AAAAATTCTC 480

183

184 TTAATAAAAG TATTGTATTA AATACACAAA GTTTTACTTT TGAAGAAGTA GAATATTTAC 540

185

186 TTAAAGGTTT AAGAAATAAA TTTCAATTAA ATTGTTATGT TAAAATTAAT AAAAATAAAC 600

187

188 CAATTATTTA TATTGATTCT ATGAGTTATC TGATTTTTTA TAATTTAATT AAACCTTATT 660

189

190 TAATTCCTCA AATGATGTAT AAAGTGCCTA ATACTATTTT ATCCGAAACT TTTTAAAAAT 720

191

192 AA 722

193

194 (2) INFORMATION FOR SEQ ID NO:4:

195

196 (i) SEQUENCE CHARACTERISTICS:

197 (A) LENGTH: 235 amino acids

198 (B) TYPE: amino acid

199 (D) TOPOLOGY: linear

200

201 (ii) MOLECULE TYPE: peptide

202

203

204

205 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

RAW SEQUENCE LISTING PATENT APPLICATION US/09/836,169

DATE: 11/20/2001
TIME: 19:31:13

INPUT SET: S36674.raw

```

206
207   Met Lys Asn Ile Lys Lys Asn Gln Val Met Asn Leu Gly Pro Asn Ser
208   1                               5                               10                               15
209
210   Lys Leu Leu Lys Glu Tyr Lys Ser Gln Leu Ile Glu Leu Asn Ile Glu
211                               20                               25                               30
212
213   Gln Phe Glu Ala Gly Ile Gly Leu Ile Leu Gly Asp Ala Tyr Ile Arg
214                               35                               40                               45
215
216   Ser Arg Asp Glu Gly Lys Thr Tyr Cys Met Gln Phe Glu Trp Lys Asn
217                               50                               55                               60
218
219   Lys Ala Tyr Met Asp His Val Cys Leu Leu Tyr Asp Gln Trp Val Leu
220   65                               70                               75                               80
221
222   Ser Pro Pro His Lys Lys Glu Arg Val Asn His Leu Gly Asn Leu Val
223                               85                               90                               95
224
225   Ile Thr Trp Gly Ala Gln Thr Phe Lys His Gln Ala Phe Asn Lys Leu
226                               100                              105                              110
227
228   Ala Asn Leu Phe Ile Val Asn Asn Lys Lys Leu Ile Pro Asn Asn Leu
229                               115                              120                              125
230
231   Val Glu Asn Tyr Leu Thr Pro Met Ser Leu Ala Tyr Trp Phe Met Asp
232   130                              135                              140
233
234   Asp Gly Gly Lys Trp Asp Tyr Asn Lys Asn Ser Leu Asn Lys Ser Ile
235   145                              150                              155                              160
236
237   Val Leu Asn Thr Gln Ser Phe Thr Phe Glu Glu Val Cys Tyr Leu Val
238                               165                              170                              175
239
240   Lys Gly Leu Arg Asn Lys Phe Gln Leu Asn Cys Tyr Val Lys Ile Asn
241                               180                              185                              190
242
243   Lys Asn Lys Pro Ile Ile Tyr Ile Asp Ser Met Ser Tyr Leu Ile Phe
244                               195                              200                              205
245
246   Tyr Asn Ile Ile Lys Pro Tyr Leu Ile Pro Gln Met Met Tyr Lys Leu
247   210                              215                              220
248
249   Pro Asn Thr Ile Ser Ser Glu Thr Phe Leu Lys
250   225                              230                              235
251

```

(2) INFORMATION FOR SEQ ID NO:5:

```

252
253
254   (i) SEQUENCE CHARACTERISTICS:
255       (A) LENGTH: 754 base pairs
256       (B) TYPE: nucleic acid
257       (C) STRANDEDNESS: single
258       (D) TOPOLOGY: linear

```

PAGE: 1

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION US/09/836,169

DATE: 11/20/2001
TIME: 19:31:13

INPUT SET: S36674.raw

Line	Error	Original Text
------	-------	---------------

PAGE: 1

SEQUENCE MISSING ITEM REPORT
PATENT APPLICATION US/09/836,169

DATE: 11/20/2001
TIME: 19:31:13

INPUT SET: S36674.raw

< < THERE ARE NO ITEMS MISSING > >

PAGE: 1

SEQUENCE CORRECTION REPORT
PATENT APPLICATION US/09/836,169

DATE: 11/20/2001
TIME: 19:31:13

INPUT SET: S36674.raw

Line

Original Text

Corrected Text